

# THE LANCET

## Global Health

### Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Rogawski ET, Liu J, Platts-Mills JA, et al, and The MAL-ED Network Investigators. Use of quantitative molecular diagnostic methods to investigate the effect of enteropathogen infections on linear growth in children in low-resource settings: longitudinal analysis of results from the MAL-ED cohort study. *Lancet Glob Health* 2018; published online Oct 1. [http://dx.doi.org/10.1016/S2214-109X\(18\)30351-6](http://dx.doi.org/10.1016/S2214-109X(18)30351-6).

## Supplementary Material

Additional MAL-ED Investigators .....	2
Methods .....	3
<b>Table S1.</b> Real time PCR assays on TaqMan Array Card used for MAL-ED diarrhoea aetiology and growth analysis.....	3
Longitudinal model methods and validation .....	4
<b>Figure S1.</b> Schematic diagram of the outcome models for the parametric g-formula. ....	5
<b>Figure S2.</b> Validation of parametric g-formula models for pathogen exposures. ....	6
<b>Figure S3.</b> Validation of parametric g-formula models for time-varying covariates and LAZ outcomes. ....	7
Enteropathogen prevalence and quantity .....	8
<b>Table S2.</b> Enteropathogen prevalence in non-diarrhoeal stools during the first 2 years of life by site among 1469 children in the MAL-ED cohort. ....	8
<b>Table S3.</b> Enteropathogen quantity in positive non-diarrhoeal stools during the first 2 years of life by site among 1469 children in the MAL-ED cohort. ....	9
<b>Table S4.</b> Enteropathogen prevalence per child in 6 month intervals during the first 2 years of life among 1469 children in the MAL-ED cohort. ....	10
<b>Table S5.</b> Enteropathogen quantity per child during the first 2 years of life among 1469 children in the MAL-ED cohort. ....	11
<b>Figure S4.</b> Associations of pathogen-attributable diarrhoea with length-for-age z-score.....	12
Site-specific effects.....	13
<b>Table S6.</b> Site-specific associations of pathogen-attributable diarrhoea episodes on LAZ 3 months following the episode .....	13
<b>Table S7.</b> Site-specific differences in length-for-age z-score at 2 years of age associated with the difference between high (site-specific 90 <sup>th</sup> percentile) and low (site-specific 10 <sup>th</sup> percentile) exposure to each of the 13 most prevalent enteropathogens in non-diarrhoeal stools among 1469 children in the MAL-ED cohort.....	14
Height attainment model sensitivity analyses .....	15
<b>Figure S5.</b> Height attainment model: qPCR vs. conventional methods.....	15
<b>Figure S6.</b> Height attainment model: non-diarrhoeal stools and diarrhoeal stools.....	16
<b>Figure S7.</b> Height attainment model: weight-for-age (WAZ) and weight-for-length (WLZ) z-scores. ....	17
Longitudinal model sensitivity analyses.....	18
<b>Figure S8.</b> Longitudinal g-formula: effects scaled per one log increase in pathogen quantity.....	18
References .....	19

## Additional MAL-ED Investigators

Angel Mendez Acosta <sup>1</sup>	Karthikeyan Ramanujam <sup>4</sup>	Fahmida Tofail <sup>11</sup>
Rosa Rios de Burga <sup>1</sup>	Anuradha Rose <sup>4</sup>	Ram Krishna Chandyo <sup>12</sup>
Cesar Banda Chavez <sup>1</sup>	Reeba Roshan <sup>4</sup>	Prakash Sunder Shrestha <sup>12</sup>
Julian Torres Flores <sup>1</sup>	Srujan L Sharma <sup>4</sup>	Rita Shrestha <sup>12</sup>
Maribel Paredes Olotegui <sup>1</sup>	Shanmuga Sundaram E <sup>4</sup>	Manjeswori Ulak <sup>12</sup>
Silvia Rengifo Pinedo <sup>1</sup>	Rahul J Thomas <sup>4</sup>	Aubrey Bauck <sup>13</sup>
Dixner Rengifo Trigos <sup>1</sup>	William K Pan <sup>5,6</sup>	Robert Black <sup>13</sup>
Angel Orbe Vasquez <sup>1</sup>	Ramya Ambikapathi <sup>6,15</sup>	Laura Caulfield <sup>13</sup>
Imran Ahmed <sup>2</sup>	J Daniel Carreon <sup>6</sup>	William Checkley <sup>13,6</sup>
Didar Alam <sup>2</sup>	Viyada Doan <sup>6</sup>	Gwenyth Lee <sup>13,21</sup>
Asad Ali <sup>2</sup>	Christel Hoest <sup>6</sup>	Kerry Schulze <sup>13</sup>
Muneera Rasheed <sup>2</sup>	Stacey Knobler <sup>6</sup>	Samuel Scott <sup>14</sup>
Sajid Soofi <sup>2</sup>	Benjamin JJ McCormick <sup>6</sup>	Laura E. Murray-Kolb <sup>16</sup>
Ali Turab <sup>2</sup>	Mark A Miller <sup>6</sup>	A Catharine Ross <sup>16</sup>
Aisha Yousafzai <sup>2</sup>	Stephanie Psaki <sup>6</sup>	Barbara Schaefer <sup>16,6</sup>
Anita KM Zaidi <sup>2</sup>	Zeba Rasmussen <sup>6</sup>	Suzanne Simons <sup>16</sup>
Binob Shrestha <sup>3</sup>	Stephanie A Richard <sup>6</sup>	Laura Pendergast <sup>17</sup>
Bishnu Bahadur Rayamajhi <sup>3</sup>	Karen H Tountas <sup>7</sup>	Cláudia B Abreu <sup>18</sup>
Tor Strand <sup>3,19</sup>	Erling Svensen <sup>8,9</sup>	Hilda Costa <sup>18</sup>
Geetha Ammu <sup>4</sup>	Caroline Amour <sup>9</sup>	Alessandra Di Moura <sup>18</sup>
Sudhir Babji <sup>4</sup>	Eliwaza Bayyo <sup>9</sup>	José Quirino Filho <sup>18,7</sup>
Anuradha Bose <sup>4</sup>	Regisiana Mvungi <sup>9</sup>	Álvaro M Leite <sup>18</sup>
Ajila T George <sup>4</sup>	John Pascal <sup>9</sup>	Noélia L Lima <sup>18</sup>
Dinesh Hariraju <sup>4</sup>	Ladislau Yarrot <sup>9</sup>	Ila F Lima <sup>18</sup>
M. Steffi Jennifer <sup>4</sup>	Leah Barrett <sup>10</sup>	Bruna LL Maciel <sup>18</sup>
Sushil John <sup>4</sup>	Rebecca Dillingham <sup>10</sup>	Pedro HQS Medeiros <sup>18</sup>
Shiny Kaki <sup>4</sup>	William A Petri, Jr <sup>10</sup>	Milena Moraes <sup>18</sup>
Priyadarshani Karunakaran <sup>4</sup>	Rebecca Scharf <sup>10</sup>	Francisco S Mota <sup>18</sup>
Beena Koshy <sup>4</sup>	AM Shamsir Ahmed <sup>11</sup>	Reinaldo B Oriá <sup>18</sup>
Robin P Lazarus <sup>4</sup>	Md Ashrafal Alam <sup>11</sup>	Josiane Quetz <sup>18</sup>
Jayaprakash Muliyil <sup>4</sup>	Umma Haque <sup>11</sup>	Alberto M Soares <sup>18</sup>
Preethi Ragasudha <sup>4</sup>	Md Iqbal Hossain <sup>11</sup>	Rosa MS Mota <sup>18</sup>
Mohan Venkata Raghava <sup>4</sup>	Munirul Islam <sup>11</sup>	Crystal L Patil <sup>20</sup>
Sophy Raju <sup>4</sup>	Mustafa Mahfuz <sup>11</sup>	Cloupas Mahopo <sup>22</sup>
Anup Ramachandran <sup>4</sup>	Dinesh Mondal <sup>11</sup>	Angelina Maphula <sup>22</sup>
Rakhi Ramadas <sup>4</sup>	Baitun Nahar <sup>11</sup>	Emanuel Nyathi <sup>22</sup>

<sup>1</sup>A.B. PRISMA, Iquitos, Peru

<sup>2</sup>Aga Khan University, Karachi, Pakistan

<sup>3</sup>Walter Reed/AFRIMS Research Unit, Kathmandu, Nepal

<sup>4</sup>Christian Medical College, Vellore, India

<sup>5</sup>Duke University, Durham, NC, USA

<sup>6</sup>Fogarty International Center/National Institutes of Health, Bethesda, MD, USA

<sup>7</sup>Foundation for the National Institutes of Health, Bethesda, MD, USA

<sup>8</sup>Haukeland University Hospital, Bergen, Norway

<sup>9</sup>Haydom Lutheran Hospital, Haydom, Tanzania

<sup>10</sup>University of Virginia, Charlottesville, VA, USA

<sup>11</sup>icddr,b, Dhaka, Bangladesh

<sup>12</sup>Institute of Medicine, Tribhuvan University, Kathmandu, Nepal

<sup>13</sup>Johns Hopkins University, Baltimore, MD, USA

<sup>14</sup>Poverty Health and Nutrition Division, International Food Policy Research Institute, Washington DC, USA

<sup>15</sup>Purdue University, Department of Nutrition Science, IN, USA

<sup>16</sup>The Pennsylvania State University, University Park, PA, USA

<sup>17</sup>Temple University, Philadelphia, PA, USA

<sup>18</sup>Universidade Federal do Ceara, Fortaleza, Brazil

<sup>19</sup>University of Bergen, Norway

<sup>20</sup>University of Illinois at Chicago, IL, USA

<sup>21</sup>University of Michigan, Department of Epidemiology, MI, USA

<sup>22</sup>University of Venda, Thohoyandou, South Africa

## Methods

**Table S1.** Real time PCR assays on TaqMan Array Card used for MAL-ED diarrhoea aetiology and growth analysis.

All the assays have been described previously and extensively validated.<sup>1–3</sup> Nucleic acid was extracted with the QIAamp Fast DNA Stool mini kit (Qiagen, Hilden, Germany) with pre-treatment steps that included bead beating. AgPath One Step RT-PCR reagents were used for qPCR reactions, which were performed on ViiA 7, or QuantStudio 7, or QuantStudio 12K Flex systems. Quantification cycles (Cqs) are the PCR cycle values at which fluorescence from amplification exceeds the background, which acts as an inverse metric of quantity of nucleic acid. Valid results required proper functioning of controls (the negative results of a sample are valid only when its external control MS2 is positive,  $Cq \leq 35$ ; the positive results are valid only when the corresponding extraction blank is negative for the relevant targets,  $Cq > 35$ ), and excluded data flagged by the real time PCR software, i.e. BADROX in combination with NOISE or SPIKE.

	Pathogen	Gene	Included in diarrhoea-growth analysis	Included in infection-growth analysis	Assay in the original microbiologic work-up (conventional methods)*
Viruses	Adenovirus 40/41	Fiber gene	✓	✓	EIA <sup>†</sup>
	Astrovirus	Capsid	✓	✓	EIA
	Norovirus GI/GII	GI ORF1-2 and GII ORF1-2	✓	✓	PCR <sup>§</sup>
	Rotavirus	<i>NSP3</i>	✓		EIA
	Sapovirus	<i>RdRp</i>	✓	✓	Not tested
Bacteria	EAEC**	<i>aaiC</i> , <i>aatA</i> , <i>aggR</i>	✓	✓	Culture/PCR
	Atypical EPEC**	<i>eae</i>	✓	✓	Culture/PCR
	Typical EPEC**	<i>bfpA</i>	✓	✓	Culture/PCR
	ETEC**	<i>LT</i> , <i>STh</i> and <i>STp</i>	✓	✓	Culture/PCR
	STEC**	<i>stx1</i> , <i>stx2</i>	✓		Culture/PCR
	<i>Aeromonas</i>	Aerolysin	✓		Culture
	<i>Campylobacter</i> spp.	<i>cadF</i> ( <i>C. jejuni/coli</i> ) and <i>cpn60</i> ( <i>Campylobacter</i> spp)	✓ ( <i>C. jejuni/coli</i> )	✓ ( <i>Campylobacter</i> spp)	EIA <sup>††</sup>
	<i>Helicobacter pylori</i>	<i>ureC</i>	✓		Not tested
	<i>Plesiomonas shigelloides</i>	<i>gyrB</i>	✓		Culture
	<i>Salmonella</i>	<i>ttr</i>	✓		Culture
Fungi	<i>Shigella/EIEC</i>	<i>ipaH</i>	✓	✓	Culture
	<i>Vibrio cholerae</i>	<i>hlyA</i>	✓		Culture
	<i>Enterocytozoon bieneusi</i>	<i>ITS</i>	✓		Not tested
	<i>Encephalitozoon intestinalis</i>	SSU rRNA	✓		Not tested
Protozoa	<i>Cryptosporidium</i>	18S rRNA	✓	✓	EIA
	<i>Cyclospora cayetanensis</i>	18S rRNA	✓		Microscopy
	<i>Cystoisospora belli</i>	18S rRNA	✓		Microscopy
	<i>Entamoeba histolytica</i>	18S rRNA	✓		EIA
	<i>Giardia</i>	18S rRNA	✓	✓	EIA
Helminth	<i>Ancylostoma duodenale</i>	<i>ITS2</i>	✓		Microscopy
	<i>Ascaris lumbricoides</i>	<i>ITS1</i>			Microscopy
	<i>Necator americanus</i>	<i>ITS2</i>			Microscopy
	<i>Strongyloides stercoralis</i>	Dispersed repetitive sequence	✓		Microscopy
	<i>Trichuris trichiura</i>	18S rRNA	✓		Microscopy
Controls	MS2	<i>MS2g1</i>			N/A
	PhHV	<i>gB</i>			N/A

\* Conventional methods were conducted in non-diarrhoeal stools monthly in the first year of life, but only quarterly in the second year of life (qPCR was conducted in non-diarrhoeal stools monthly for both years).

<sup>†</sup> Pan-adenovirus EIA only.

<sup>§</sup> Norovirus PCR was only performed on diarrhoeal and non-diarrhoeal stools from a randomly selected 10% of participants.

\*\**E. coli* pathotypes were defined as follows: EAEC (*aaiC*, or *aatA*, or both), atypical EPEC (*eae* without *bfpA*, *stx1*, and *stx2*), typical EPEC (*bfpA*), ETEC (*STh*, *STp*, or *LT*), STEC (*eae* without *bfpA* and with *stx1*, *stx2*, or both). EAEC was alternatively defined by *aggR* in sensitivity analysis.

<sup>††</sup>A single EIA was used, which has been shown to detect some *Campylobacter* species other than *C. jejuni* and *C. coli*.<sup>4</sup>

## Longitudinal model methods and validation

All analyses were conducted in SAS version 9.4 (Cary, North Carolina, USA). To implement the parametric g-formula, we first used the observed data to estimate  $\beta$ -coefficients in a series of longitudinal repeated measures models for each time-varying covariate in the 6-month intervals. Results from models with 3-month and 6-month intervals were largely consistent; we present the 6-month interval results because 3-month interval-based estimates were less precise and more sensitive to model specification. All models included an indicator for the interval, LAZ at the beginning of the interval, site, sex, and SES as covariates. The time-varying covariates modelled were:

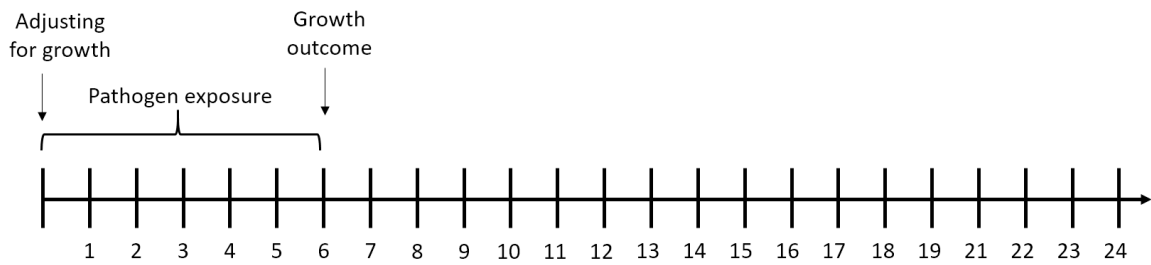
1. Percent days with exclusive breastfeeding: modelled with Poisson regression with an offset for the number of days under surveillance in the period, and also adjusting for maternal age.
2. Incidence of diarrhoea (number of episodes): modelled with Poisson regression and also adjusting for exclusive breastfeeding, and detection of *Campylobacter*, *Shigella*, and tEPEC in the previous interval.
3. Macrolide use (yes/no): modelled with logistic regression and also adjusting for exposure to EAEC, *Campylobacter*, *Giardia*, *Shigella*, and tEPEC in the previous interval, number of diarrhoea episodes, exclusive breastfeeding, and improved sanitation. Macrolide use was included since it was a strong predictor of *Campylobacter*, *Shigella*, and tEPEC infection.
4. The number of detections of pathogens that were not the exposure of interest (*Campylobacter* spp., *Giardia*, EAEC, tEPEC, and *Shigella*): modelled with Poisson regression with an offset for the number of stool samples tested for that pathogen, and also adjusting for exposure to that pathogen in the previous interval, exposure to the other four pathogens in the previous interval, exclusive breastfeeding, number of diarrhoea episodes, macrolide use, maternal height, improved sanitation, and improved drinking water.
5. The average quantity of these pathogens: modelled with the same covariates using log-linear regression only among those with positive detections.
6. The number of detections or average quantity of the pathogen of interest: similarly modelled with Poisson and log-linear regression, respectively, but further adjusted for exposure to the other pathogens (from step 4) in the current interval.
7. The outcome, LAZ, at the end of the interval: modelled using a separate linear regression model for each 6-month time period, and additionally adjusted for the current and lagged pathogen exposure variables (no lags during the first 6-month period, and progressively more lags for later periods), exclusive breastfeeding (first period only), and maternal height (Figure S1).

These time-varying covariates were considered for inclusion based on a causal directed acyclic graph.<sup>4</sup> Covariates were dropped from models for specific pathogen detections if they did not have a strong ( $|\text{LAZ difference}| > 0.02$ ) or significant association ( $p < 0.05$ ) with that pathogen and if model fit improved by Akaike information criterion (AIC).

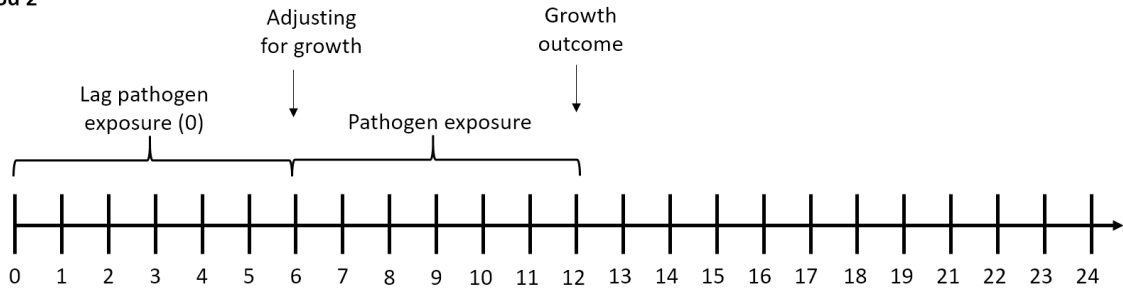
We then used Monte Carlo simulations with the estimated coefficients from the models above to predict the time-varying covariates, pathogen exposures, and LAZ outcomes for each interval to 2 years of age in a random sample of 50,000 replicates from the study population at baseline. The simulation was completed for the natural course scenario, in which all pathogen exposures were predicted based on the observed data, to validate model fit. The simulated data under the natural course scenario fit the observed data well, such that the interval-specific distributions of the time-varying covariates and LAZ outcomes were similar between the observed and simulated data (Figure S2).

The simulations were then completed for high and low pathogen exposure scenarios, which produced the predicted covariates and outcomes had all children been exposed to high pathogen exposure, defined by the 90<sup>th</sup> percentile in each interval, versus had all children been exposed to low pathogen exposure, defined by the 10<sup>th</sup> percentile in each interval. To estimate the cumulative effect of pathogen exposure on LAZ at 2 years of age, we averaged the predicted LAZ outcomes at 2 years across individuals in the high and low pathogen exposure scenarios and took their difference to estimate the population-standardized LAZ difference. A similar process was used to estimate the cumulative effect of pathogen exposure on LAZ at 2 years of age per one log increase in pathogen quantity. Confidence intervals were constructed by bootstrap of the above steps with 1000 replicates.

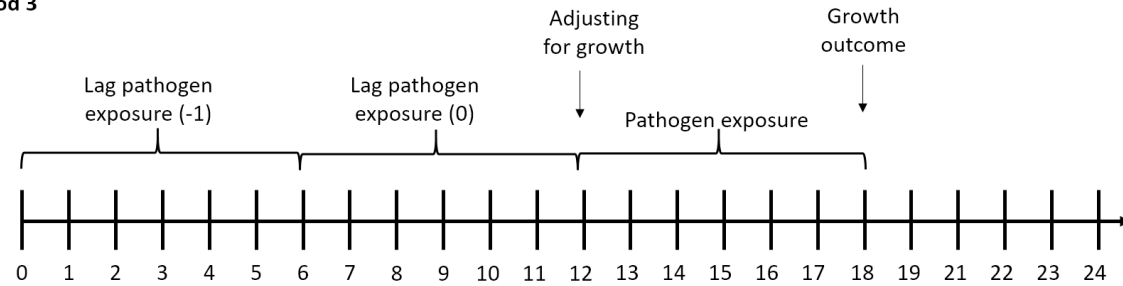
**Period 1**



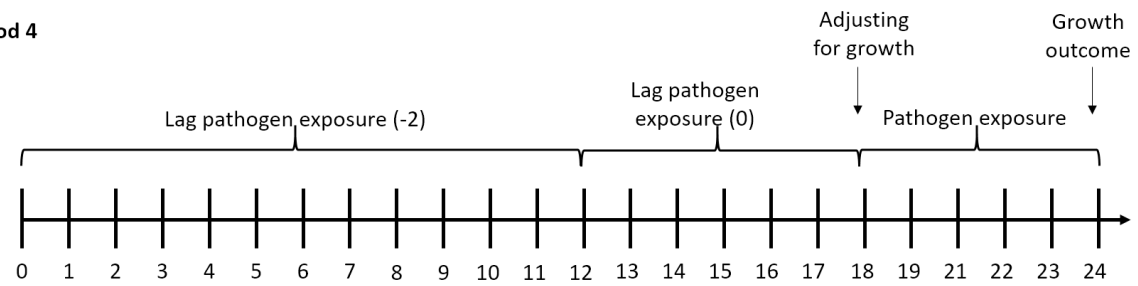
**Period 2**



**Period 3**

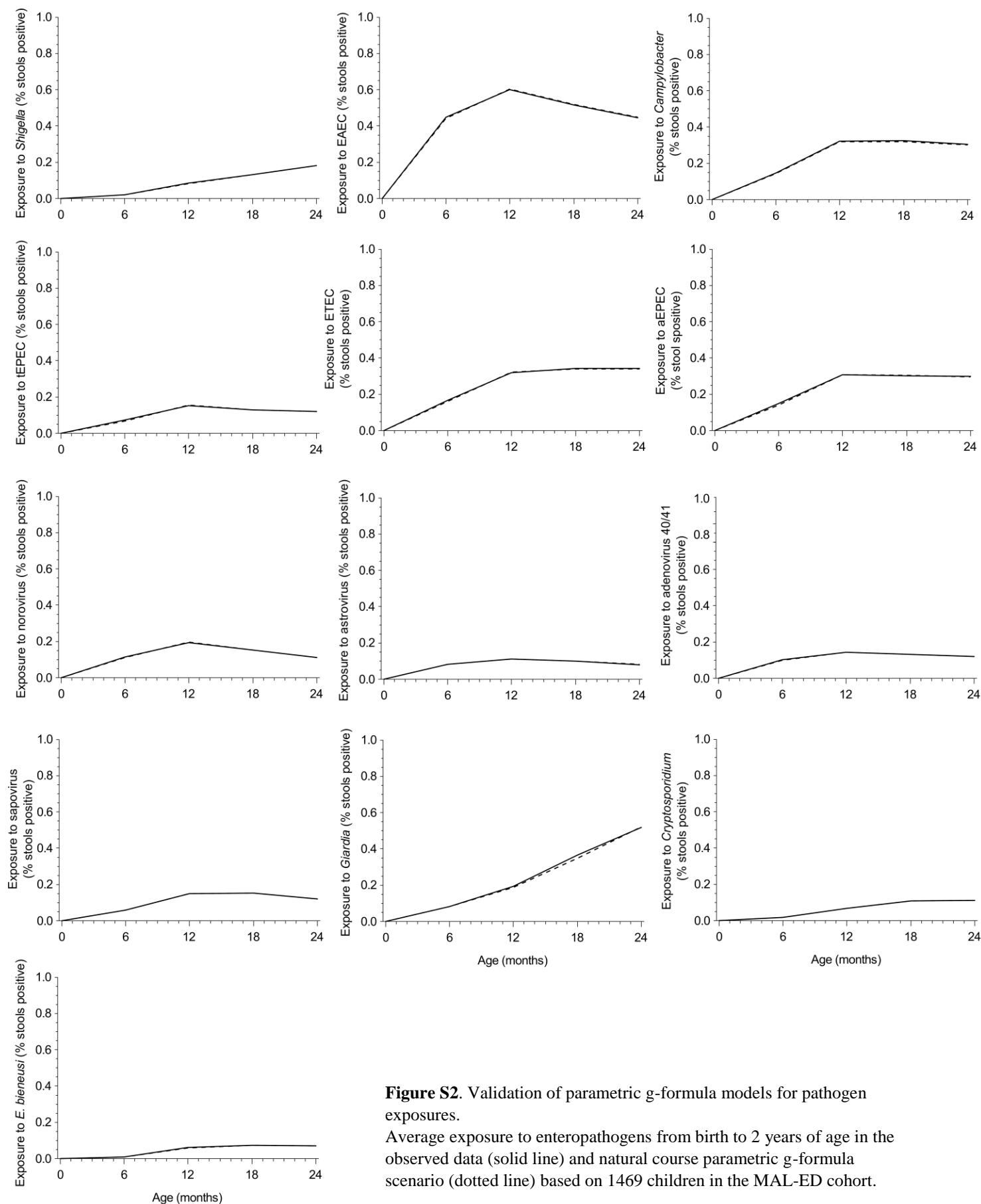


**Period 4**



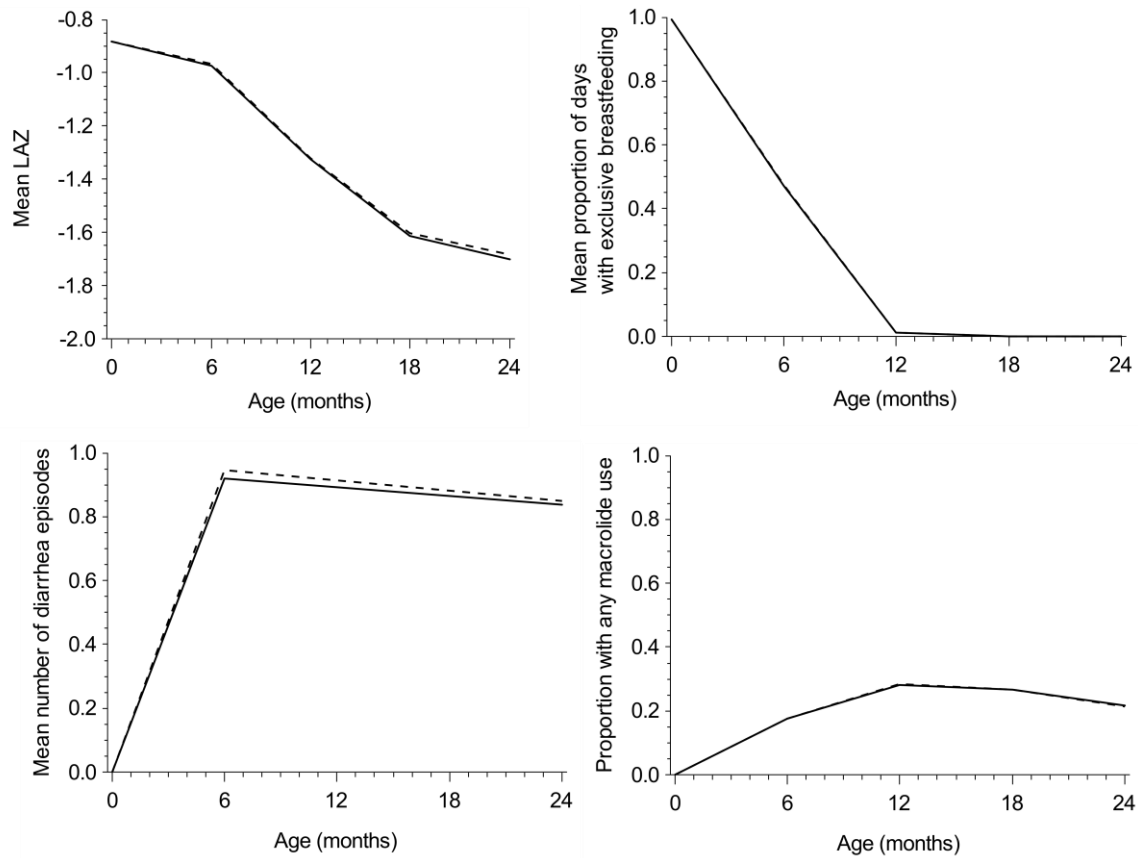
Month of age

**Figure S1.** Schematic diagram of the outcome models for the parametric g-formula.



**Figure S2.** Validation of parametric g-formula models for pathogen exposures.

Average exposure to enteropathogens from birth to 2 years of age in the observed data (solid line) and natural course parametric g-formula scenario (dotted line) based on 1469 children in the MAL-ED cohort.



**Figure S3.** Validation of parametric g-formula models for time-varying covariates and LAZ outcomes. Average LAZ, proportion of days with exclusive breastfeeding, number of diarrhoea episodes, and macrolide use from birth to 2 years of age in the observed data (solid line) and natural course parametric g-formula scenario (dotted line) based on 1469 children in the MAL-ED cohort.



## Enteropathogen prevalence and quantity

**Table S2.** Enteropathogen prevalence in non-diarrhoeal stools during the first 2 years of life by site among 1469 children in the MAL-ED cohort.

		Percent non-diarrhoeal stools positive							
		N (%)							
	Site	All	Dhaka, Bangladesh	Vellore, India	Bhaktapur, Nepal	Fortaleza, Brazil	Loreto, Peru	Venda, South Africa	Haydom, Tanzania
Bacteria	<i>Shigella</i>	3237 (10.8)	564 (13.1)	592 (12.4)	290 (5.8)	139 (4.9)	574 (13.6)	321 (7.0)	757 (17.8)
	EAEC	15314 (51.0)	1877 (43.6)	3052 (63.9)	2503 (49.6)	857 (30.1)	2621 (62.2)	1654 (35.9)	2750 (65.0)
	<i>Campylobacter</i>	8398 (28.0)	1702 (39.4)	1177 (24.7)	1325 (26.3)	428 (15.0)	1044 (24.8)	525 (11.4)	2197 (51.8)
	tEPEC	3653 (12.2)	862 (20.0)	770 (16.2)	344 (6.8)	97 (3.4)	590 (14.0)	201 (4.4)	789 (18.6)
	ETEC	9056 (30.0)	2026 (46.8)	1471 (30.7)	1057 (20.9)	222 (7.7)	1349 (32.0)	578 (12.6)	2353 (54.8)
	aEPEC	8054 (26.8)	1083 (25.1)	1339 (28.1)	1535 (30.4)	742 (25.9)	1203 (28.6)	924 (20.2)	1228 (28.8)
Viruses	Norovirus	4338 (14.5)	836 (19.4)	730 (15.4)	628 (12.4)	161 (5.7)	806 (19.1)	477 (10.5)	700 (16.5)
	Astrovirus	2862 (9.6)	792 (18.3)	492 (10.4)	253 (5.0)	45 (1.6)	695 (16.6)	313 (6.8)	272 (6.4)
	Adenovirus 40/41	3791 (12.7)	779 (18.0)	787 (16.5)	296 (5.9)	126 (4.4)	969 (23.3)	491 (10.7)	343 (8.1)
	Sapovirus	3680 (12.3)	746 (17.3)	662 (13.9)	536 (10.6)	122 (4.3)	660 (15.7)	501 (10.9)	453 (10.7)
Protozoa	<i>Giardia</i>	8898 (29.7)	923 (21.4)	1804 (37.6)	1036 (20.5)	442 (15.5)	1760 (42.3)	1099 (23.6)	1834 (43.8)
	<i>Cryptosporidium</i>	2363 (7.9)	364 (8.4)	257 (5.4)	318 (6.3)	38 (1.3)	528 (12.5)	250 (5.5)	608 (14.3)
	<i>E. bieneusi</i> *	1674 (5.6)	354 (8.2)	135 (2.8)	161 (3.2)	19 (0.7)	256 (6.1)	190 (4.2)	559 (13.2)

\**E. bieneusi* is an intracellular parasitic fungus

**Table S3.** Enteropathogen quantity in positive non-diarrhoeal stools during the first 2 years of life by site among 1469 children in the MAL-ED cohort.

		Mean quantity in log (copy number) per gram of stool (standard deviation)							
	Site	All	Dhaka, Bangladesh	Vellore, India	Bhaktapur, Nepal	Fortaleza, Brazil	Loreto, Peru	Venda, South Africa	Haydom, Tanzania
Bacteria	<i>Shigella</i>	5.62 (1.29)	5.61 (1.19)	5.47 (1.20)	5.64 (1.26)	5.14 (1.24)	5.81 (1.43)	5.05 (0.94)	5.91 (1.37)
	EAEC	6.12 (1.52)	5.97 (1.56)	6.06 (1.46)	6.59 (1.46)	5.66 (1.37)	6.48 (1.65)	5.35 (1.09)	6.14 (1.52)
	<i>Campylobacter</i>	5.74 (1.33)	5.95 (1.36)	5.37 (1.21)	6.06 (1.41)	4.88 (0.85)	5.85 (1.37)	4.91 (1.14)	5.91 (1.21)
	tEPEC	5.83 (1.43)	5.79 (1.49)	5.71 (1.37)	6.28 (1.34)	5.38 (1.18)	5.88 (1.55)	5.37 (1.08)	5.92 (1.41)
	ETEC	5.88 (1.60)	5.68 (1.56)	5.65 (1.49)	6.47 (1.64)	5.47 (1.38)	5.89 (1.73)	5.27 (1.17)	6.12 (1.59)
	aEPEC	5.45 (1.24)	5.46 (1.26)	5.35 (1.19)	5.79 (1.23)	5.30 (1.13)	5.70 (1.37)	4.88 (0.91)	5.40 (1.25)
Viruses	Norovirus	5.25 (1.10)	5.29 (1.10)	5.18 (1.03)	5.27 (1.14)	5.01 (1.00)	5.41 (1.14)	5.05 (1.03)	5.29 (1.10)
	Astrovirus	5.37 (1.71)	5.25 (1.55)	5.17 (1.55)	6.17 (1.99)	5.64 (1.93)	5.22 (1.68)	5.15 (1.55)	5.89 (2.00)
	Adenovirus 40/41	5.35 (1.64)	5.28 (1.44)	5.17 (1.47)	6.10 (2.36)	5.55 (1.61)	5.22 (1.45)	5.00 (1.35)	6.08 (2.08)
	Sapovirus	5.55 (1.50)	5.63 (1.52)	5.37 (1.44)	5.69 (1.61)	5.53 (1.40)	5.50 (1.52)	5.38 (1.39)	5.83 (1.46)
Protozoa	<i>Giardia</i>	6.36 (1.77)	5.72 (1.66)	5.98 (1.63)	7.54 (1.74)	5.92 (1.32)	6.63 (1.78)	5.40 (1.45)	6.80 (1.67)
	<i>Cryptosporidium</i>	5.41 (1.34)	5.46 (1.47)	5.21 (1.13)	5.52 (1.37)	4.95 (1.08)	5.64 (1.47)	5.06 (1.15)	5.39 (1.25)
	<i>E. bieneusi</i> *	5.63 (1.39)	5.87 (1.35)	5.68 (1.46)	5.67 (1.45)	5.45 (1.18)	5.62 (1.53)	5.37 (1.30)	5.54 (1.34)

\**E. bieneusi* is an intracellular parasitic fungus

**Table S4.** Enteropathogen prevalence per child in 6 month intervals during the first 2 years of life among 1469 children in the MAL-ED cohort.

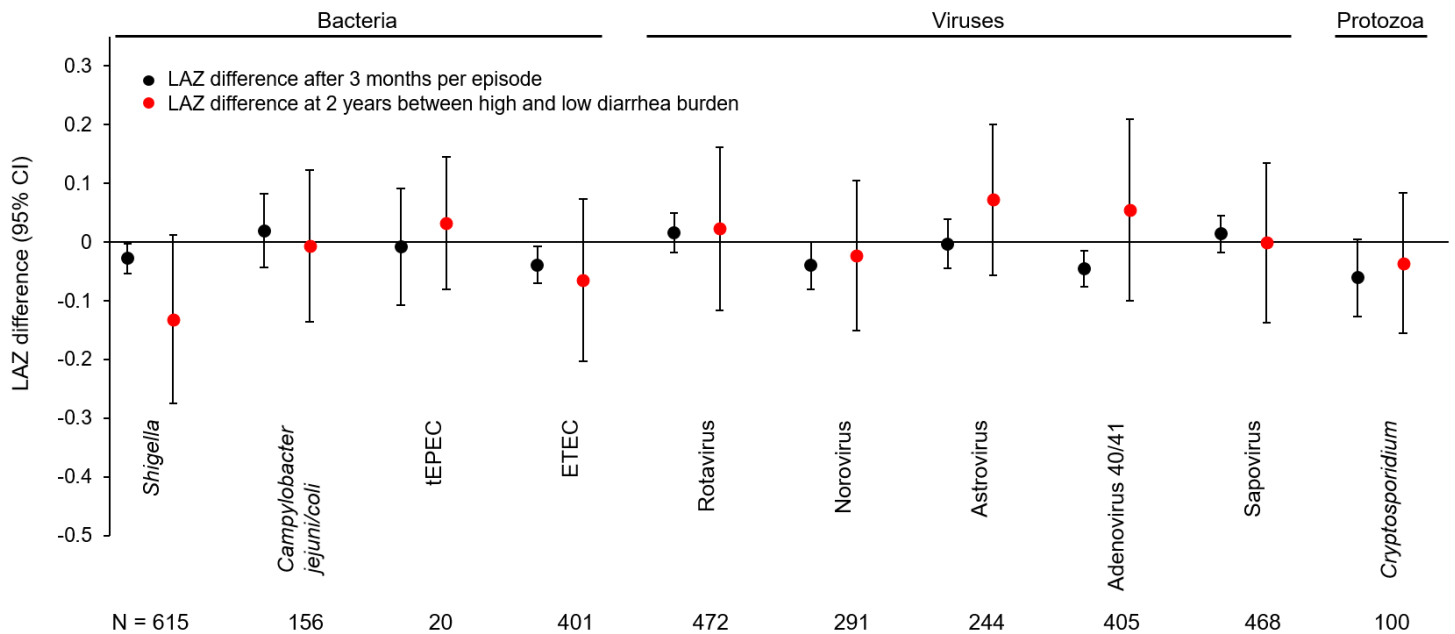
		Median percent non-diarrhoeal stools positive per child (10 <sup>th</sup> , 90 <sup>th</sup> percentiles)				
	Months	1-24	1-6	7-12	13-18	19-24
Bacteria	<i>Shigella</i>	0.09 (0–0.24)	0 (0–0)	0 (0–0.33)	0 (0–0.40)	0.17 (0–0.50)
	EAEC	0.50 (0.25–0.74)	0.4 (0–1)	0.6 (0.20–1)	0.5 (0.17–0.83)	0.5 (0–0.80)
	<i>Campylobacter</i>	0.24 (0.05–0.57)	0 (0–0.50)	0.33 (0–0.75)	0.25 (0–0.80)	0.2 (0–0.80)
	tEPEC	0.10 (0–0.25)	0 (0–0.25)	0.17 (0–0.40)	0 (0–0.33)	0 (0–0.33)
	ETEC	0.27 (0.06–0.57)	0.17 (0–0.5)	0.25 (0–0.75)	0.33 (0–0.75)	0.33 (0–0.75)
	aEPEC	0.26 (0.13–0.41)	0 (0–0.40)	0.33 (0–0.60)	0.33 (0–0.60)	0.33 (0–0.60)
Viruses	Norovirus	1.14 (0.04–0.26)	0 (0–0.33)	0.17 (0–0.50)	0.17 (0–0.40)	0 (0–0.33)
	Astrovirus	0.08 (0–0.22)	0 (0–0.33)	0 (0–0.33)	0 (0–0.33)	0 (0–0.25)
	Adenovirus 40/41	0.10 (0–0.27)	0 (0–0.33)	0 (0–0.40)	0 (0–0.33)	0 (0–0.33)
	Sapovirus	0.11 (0–0.23)	0 (0–0.20)	0.17 (0–0.40)	0.17 (0–0.40)	0 (0–0.33)
Protozoa	<i>Giardia</i>	0.25 (0–0.63)	0 (0–0.33)	0 (0–0.67)	0.2 (0–1)	0.6 (0–1)
	<i>Cryptosporidium</i>	0.05 (0–0.18)	0 (0–0)	0 (0–0.25)	0 (0–0.33)	0 (0–0.33)
	<i>E. bienewsi</i> *	0 (0–0.16)	0 (0–0)	0 (0–0.33)	0 (0–0.33)	0 (0–0.33)

\**E. bienewsi* is an intracellular parasitic fungus

**Table S5.** Enteropathogen quantity per child during the first 2 years of life among 1469 children in the MAL-ED cohort.

		Median quantity in log (copy number) per gram of stool (10 <sup>th</sup> , 90 <sup>th</sup> percentiles)				
	Months	1-24	1-6	7-12	13-18	19-24
Bacteria	<i>Shigella</i>	0.49 (0–1.39)	0 (0–0)	0 (0–1.85)	0 (0–2.27)	0.74 (0–3.02)
	EAEC	3.12 (1.37–4.69)	2.71 (0–6.28)	3.74 (0.98–6.39)	3.01 (0.82–5.44)	2.46 (0–4.71)
	<i>Campylobacter</i>	1.32 (0.27–3.30)	0 (0–2.78)	1.60 (0–4.35)	1.39 (0–4.52)	1.12 (0–4.21)
	tEPEC	0.62 (0–1.46)	0 (0–1.75)	0.72 (0–2.50)	0 (0–2.18)	0 (0–1.99)
	ETEC	1.63 (0.31–3.38)	0 (0–0)	1.56 (0–3.51)	1.52 (0–3.30)	1.39 (0–3.34)
	aEPEC	1.39 (0.63–2.30)	0 (0–1.32)	0.70 (0–2.23)	0.75 (0–2.17)	0 (0–1.89)
Viruses	Norovirus	0.71 (0.18–1.39)	0 (0–2.05)	0.92 (0–2.52)	0.75 (0–1.96)	0 (0–1.56)
	Astrovirus	0.43 (0–1.17)	0 (0–1.50)	0 (0–1.86)	0 (0–1.74)	0 (0–1.49)
	Adenovirus 40/41	0.56 (0–1.44)	0 (0–1.72)	0 (0–2.22)	0 (0–2.03)	0 (0–1.92)
	Sapovirus	0.62 (0–1.27)	0 (0–0)	0 (0–1.48)	0 (0–1.94)	0 (0–1.85)
Protozoa	<i>Giardia</i>	1.45 (0–4.28)	0 (0–1.40)	0 (0–3.95)	1.22 (0–6.77)	3.38 (0–7.40)
	<i>Cryptosporidium</i>	0.32 (0–0.99)	0 (0–1.84)	0.70 (0–2.57)	0.72 (0–2.41)	0.72 (0–2.16)
	<i>E. bienewsi</i> *	0 (0–0.89)	0 (0–0)	0 (0–1.65)	0 (0–1.59)	0 (0–1.66)

\**E. bienewsi* is an intracellular parasitic fungus



**Figure S4.** Associations of pathogen-attributable diarrhoea with length-for-age z-score.

Per-episode association of diarrhoea with length-for-age z-score (LAZ) 3 months after the episode (LAZ difference and 95% CI; black dots) and LAZ difference and 95% CI at 2 years of age between high (90<sup>th</sup> percentile) and low (10<sup>th</sup> percentile) pathogen-attributable diarrhoea burden (red dots). Analysis includes 37951 observed child-months among 1469 children in the MAL-ED cohort with molecular testing of stool samples. Per-episode estimates were adjusted for age, site, sex, SES, maternal height, LAZ at the beginning of the interval, exclusive breastfeeding, and non-attributable diarrhoea episodes in the same period. Two-year estimates were adjusted for site, sex, SES, maternal height, enrolment LAZ, exclusive breastfeeding in the first 6 months of life, number of antibiotic courses, and number of non-attributable diarrhoea episodes.

## Site-specific effects

**Table S6.** Site-specific associations of pathogen-attributable diarrhoea episodes on LAZ 3 months following the episode adjusted for age, sex, SES, maternal height, and antibiotic treatment of the episode among 1469 children in the MAL-ED cohort.

Site	Diarrhoea aetiology			
	Any infectious aetiology	Bacterial*	Viral†	Parasitic‡
Dhaka, Bangladesh	-0.06 (-0.08 to -0.03)	-0.05 (-0.08 to -0.02)	-0.03 (-0.06–0.00)	0.01 (-0.10–0.11)
Vellore, India	-0.04 (-0.07 to -0.01)	-0.02 (-0.06–0.02)	-0.03 (-0.07–0.01)	-0.04 (-0.17–0.10)
Bhaktapur, Nepal	0.02 (-0.01–0.05)	0.00 (-0.05–0.05)	0.05 (0.01–0.08)	-0.14 (-0.26 to -0.02)
Fortaleza, Brazil	-0.10 (-0.23–0.02)	-0.03 (-0.19–0.12)	-0.10 (-0.27–0.07)	-0.27 (-0.60–0.05)
Loreto, Peru	-0.02 (-0.05–0.01)	-0.05 (-0.09 to -0.01)	0.01 (-0.02–0.05)	-0.10 (-0.19–0.00)
Venda, South Africa	0.05 (-0.06–0.17)	0.03 (-0.16–0.22)	0.05 (-0.10–0.19)	N/A§
Haydom, Tanzania	-0.13 (-0.22 to -0.04)	-0.04 (-0.18–0.09)	-0.15 (-0.28 to -0.01)	-0.15 (-0.41–0.11)

\* Includes EAEC, aEPEC, tEPEC, ETEC, STEC, *Aeromonas*, *C. jejuni/coli*, *H. pylori*, *Plesiomonas*, *Salmonella*, *Shigella*, and *V. cholera*

† Includes adenovirus 40/41, astrovirus, norovirus GI/GII, rotavirus, sapovirus

‡ Includes *E. bieneusi*, *E. intestinalis*, *Cryptosporidium*, *Cyclospora*, *Isospora*, *E. histolytica*, *Giardia*, *Ancylostoma*, *Strongyloides*, and *Trichuris*

§ No parasite-attributable diarrhoea episodes were identified in Venda, South Africa

**Table S7.** Site-specific differences in length-for-age z-score at 2 years of age associated with the difference between high (site-specific 90<sup>th</sup> percentile) and low (site-specific 10<sup>th</sup> percentile) exposure to each of the 13 most prevalent enteropathogens in non-diarrhoeal stools among 1469 children in the MAL-ED cohort.

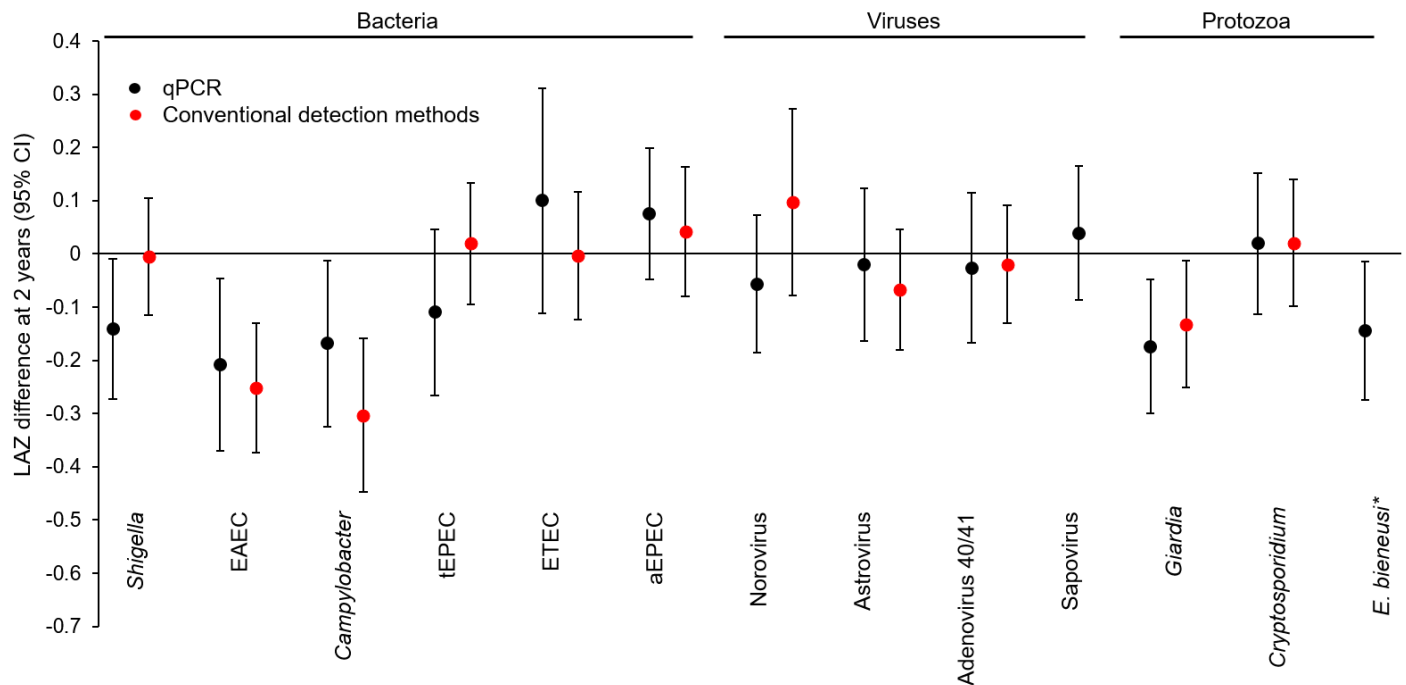
		All sites	Dhaka, Bangladesh	Vellore, India	Bhaktapur, Nepal
Bacteria	<i>Shigella</i>	-0.14 (-0.27 to -0.01)	-0.23 (-0.53–0.08)	0.00 (-0.28–0.28)	0.21 (-0.04–0.46)
	EAEC	-0.21 (-0.37 to -0.05)	-0.26 (-0.52–0.01)	0.11 (-0.16–0.39)	0.07 (-0.19–0.32)
	<i>Campylobacter</i>	-0.17 (-0.32 to -0.01)	-0.19 (-0.50–0.13)	-0.05 (-0.33–0.24)	-0.34 (-0.61 to -0.07)
	tEPEC	-0.11 (-0.27–0.05)	0.15 (-0.14–0.44)	-0.10 (-0.38–0.18)	-0.38 (-0.64 to -0.11)
	ETEC	0.10 (-0.11–0.31)	0.01 (-0.29–0.31)	-0.07 (-0.35–0.21)	0.04 (-0.23–0.30)
	aEPEC	0.08 (-0.05–0.20)	-0.03 (-0.31–0.25)	-0.18 (-0.47–0.11)	0.14 (-0.12–0.40)
Viruses	Norovirus	-0.06 (-0.19–0.07)	-0.25 (-0.53–0.02)	-0.26 (-0.53–0.01)	0.17 (-0.08–0.42)
	Astrovirus	-0.02 (-0.16–0.12)	-0.04 (-0.31–0.23)	-0.08 (-0.35–0.19)	0.02 (-0.23–0.28)
	Adenovirus 40/41	-0.03 (-0.17–0.11)	-0.08 (-0.35–0.19)	0.12 (-0.14–0.38)	0.14 (-0.10–0.38)
	Sapovirus	0.04 (-0.09–0.16)	-0.15 (-0.43–0.13)	0.18 (-0.09–0.44)	0.16 (-0.09–0.41)
Protozoa	<i>Giardia</i>	-0.17 (-0.30 to -0.05)	0.09 (-0.18–0.36)	-0.25 (-0.52–0.03)	-0.18 (-0.43–0.06)
	<i>Cryptosporidium</i>	0.02 (-0.11–0.15)	0.23 (-0.04–0.51)	-0.04 (-0.30–0.22)	-0.18 (-0.45–0.09)
	<i>E. bienewsi</i> *	-0.14 (-0.27 to -0.01)	0.09 (-0.19–0.38)	-0.05 (-0.32–0.21)	-0.15 (-0.39–0.10)

		Fortaleza, Brazil	Loreto, Peru	Venda, South Africa	Haydom, Tanzania
Bacteria	<i>Shigella</i>	-0.03 (-0.48–0.42)	-0.23 (-0.51–0.04)	-0.23 (-0.55–0.08)	-0.36 (-0.67 to -0.04)
	EAEC	-0.28 (-0.72–0.16)	-0.04 (-0.31–0.23)	-0.34 (-0.67–0.00)	-0.40 (-0.72 to -0.07)
	<i>Campylobacter</i>	-0.01 (-0.4–0.38)	0.08 (-0.19–0.34)	-0.30 (-0.61–0.01)	-0.06 (-0.37–0.24)
	tEPEC	0.00 (-0.39–0.39)	0.09 (-0.19–0.37)	-0.11 (-0.42–0.20)	-0.03 (-0.36–0.31)
	ETEC	0.14 (-0.28–0.56)	-0.17 (-0.45–0.11)	0.11 (-0.22–0.44)	0.38 (0.06–0.71)
	aEPEC	-0.17 (-0.59–0.26)	0.31 (0.03–0.60)	0.13 (-0.18–0.45)	0.31 (-0.01–0.63)
Viruses	Norovirus	0.06 (-0.35, 0.47)	-0.18 (-0.46–0.10)	0.03 (-0.29–0.35)	0.18 (-0.12–0.47)
	Astrovirus	-0.55 (-0.94 to -0.16)	-0.13 (-0.39–0.12)	0.16 (-0.14–0.47)	0.20 (-0.09–0.48)
	Adenovirus 40/41	-0.01 (-0.43–0.41)	-0.14 (-0.41–0.13)	-0.19 (-0.49–0.11)	-0.04 (-0.32–0.25)
	Sapovirus	0.33 (-0.09–0.75)	0.16 (-0.10–0.43)	0.05 (-0.26–0.36)	-0.19 (-0.48–0.10)
Protozoa	<i>Giardia</i>	0.16 (-0.28–0.61)	-0.33 (-0.60 to -0.05)	-0.35 (-0.66 to -0.04)	-0.10 (-0.40–0.21)
	<i>Cryptosporidium</i>	-0.03 (-0.43–0.37)	0.03 (-0.23–0.29)	0.38 (0.06–0.69)	-0.07 (-0.37–0.23)
	<i>E. bienewsi</i> *	-0.06 (-0.46–0.34)	-0.19 (-0.46–0.07)	-0.24 (-0.55–0.07)	-0.24 (-0.54–0.05)

\**E. bienewsi* is an intracellular parasitic fungus

## Height attainment model sensitivity analyses



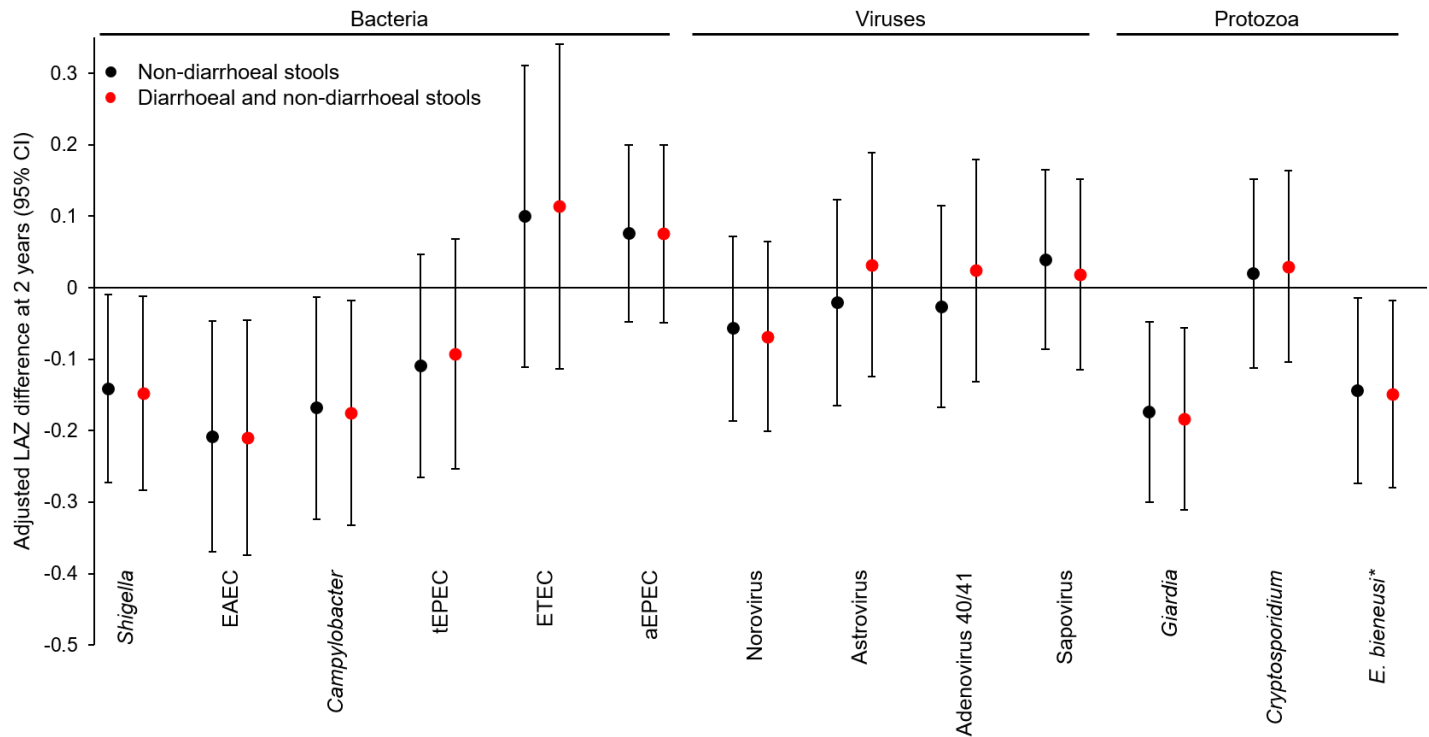
**Figure S5.** Height attainment model: qPCR vs. conventional methods.

Difference and 95% confidence interval in length-for-age z-score (LAZ) at 2 years of age associated with the difference between high (90<sup>th</sup> percentile) and low (10<sup>th</sup> percentile) pathogen prevalence for each of the 13 most prevalent enteropathogens using quantitative PCR detection methods (black) and conventional detection methods<sup>5</sup> (Table S1; red) in non-diarrhoeal stools among 1469 children in the MAL-ED cohort. Estimates are adjusted for site, enrolment LAZ, sex, SES, exclusive breastfeeding in the first 6 months of life, and maternal height.

Note: Sapovirus and *E. bieneusi* were not assayed using conventional methods and norovirus was only tested using conventional methods in random 10% of children. Conventional methods were conducted in non-diarrhoeal stools monthly in the first year of life and only quarterly in the second year, while qPCR was conducted in non-diarrhoeal stools monthly in both years.

\**E. bieneusi* is an intracellular parasitic fungus

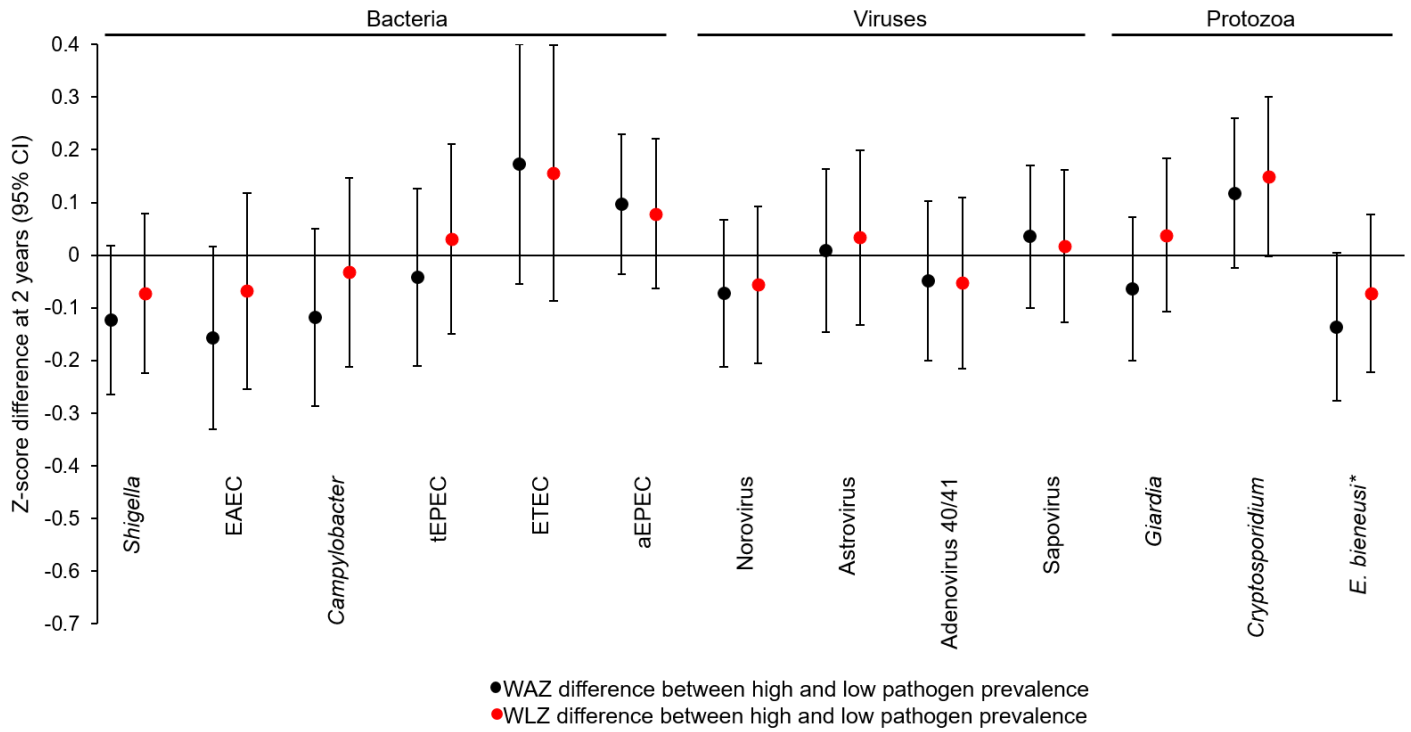




**Figure S6.** Height attainment model: non-diarrhoeal stools and diarrhoeal stools.

Difference and 95% confidence interval in LAZ at 2 years of age associated with the difference between high (90<sup>th</sup> percentile) and low (10<sup>th</sup> percentile) pathogen prevalence for each of the 13 most prevalent enteropathogens in non-diarrhoeal stools (black) and all stools (diarrhoeal and non-diarrhoeal; red) among 1469 children in the MAL-ED cohort. Estimates are adjusted for site, enrolment WAZ and LAZ, sex, SES, exclusive breastfeeding in the first 6 months of life, and maternal height.

\**E. bieneusi* is an intracellular parasitic fungus

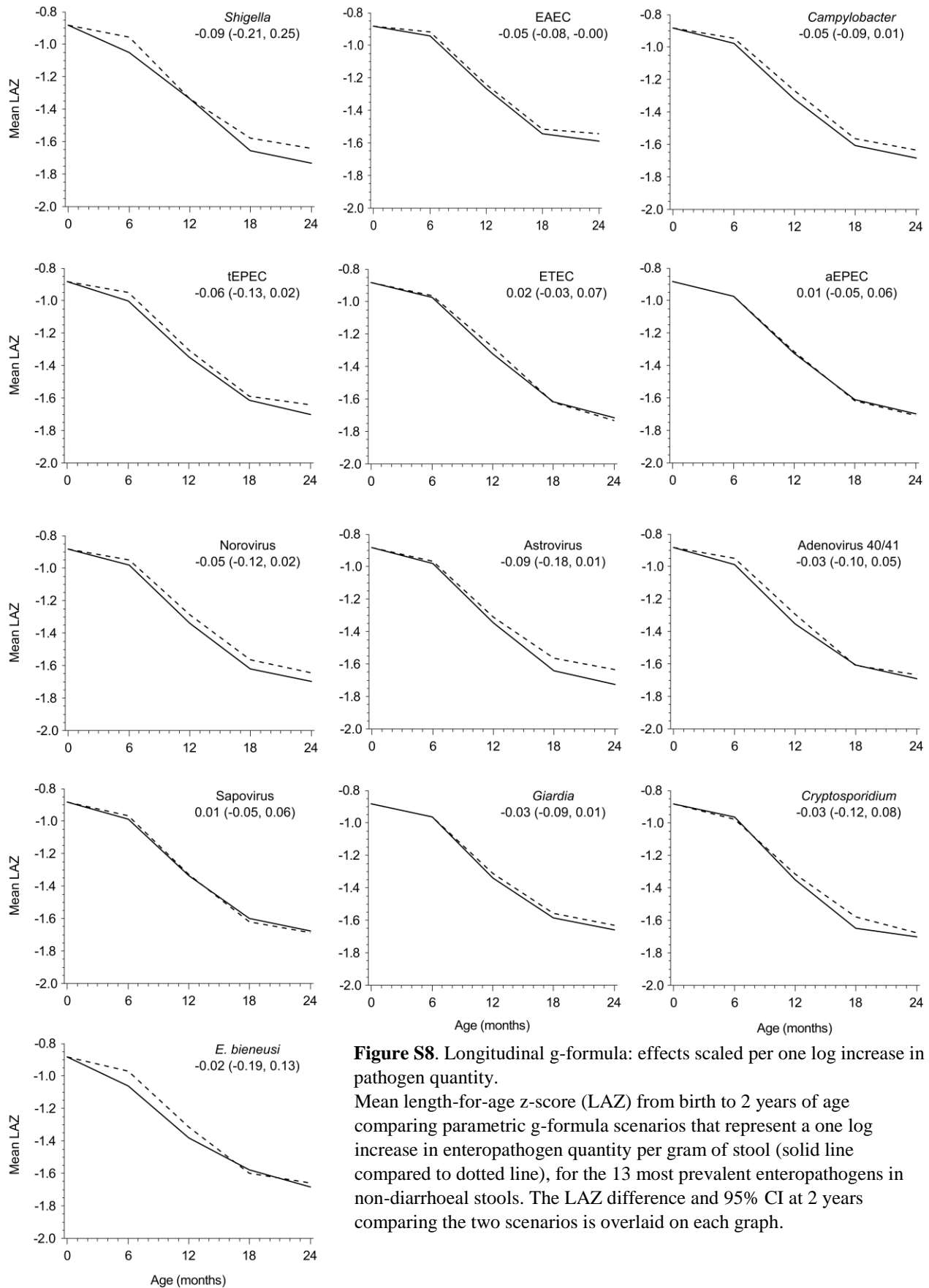


**Figure S7.** Height attainment model: weight-for-age (WAZ) and weight-for-length (WLZ) z-scores.

Difference and 95% confidence interval in WAZ and WLZ at 2 years of age associated with the difference between high (90<sup>th</sup> percentile) and low (10<sup>th</sup> percentile) pathogen prevalence for each of the 13 most prevalent enteropathogens in non-diarrhoeal stools among 1469 children in the MAL-ED cohort. Estimates are adjusted for site, enrolment WAZ and LAZ, sex, SES, exclusive breastfeeding in the first 6 months of life, and maternal height.

\**E. bieneusi* is an intracellular parasitic fungus

## Longitudinal model sensitivity analyses



**Figure S8.** Longitudinal g-formula: effects scaled per one log increase in pathogen quantity.

Mean length-for-age z-score (LAZ) from birth to 2 years of age comparing parametric g-formula scenarios that represent a one log increase in enteropathogen quantity per gram of stool (solid line compared to dotted line), for the 13 most prevalent enteropathogens in non-diarrhoeal stools. The LAZ difference and 95% CI at 2 years comparing the two scenarios is overlaid on each graph.

## References

- 1 Liu J, Gratz J, Amour C, *et al.* Optimization of Quantitative PCR Methods for Enteropathogen Detection. *PloS One* 2016; **11**: e0158199.
- 2 Liu J, Platts-Mills JA, Juma J, *et al.* Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet Lond Engl* 2016; **388**: 1291–301.
- 3 Liu J, Gratz J, Amour C, *et al.* A Laboratory-Developed TaqMan Array Card for Simultaneous Detection of 19 Enteropathogens. *J Clin Microbiol* 2013; **51**: 472–80.
- 4 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiol Camb Mass* 1999; **10**: 37–48.
- 5 Houpt E, Gratz J, Kosek M, *et al.* Microbiologic methods utilized in the MAL-ED cohort study. *Clin Infect Dis Off Publ Infect Dis Soc Am* 2014; **59 Suppl 4**: S225–32.